ENTERED   OIPE

RAW SEQUENCE LISTING   DATE: 04/23/2002
PATENT APPLICATION: US/10/040,895   TIME: 15:35:39

Input Set : A:\Tb5072.txt
Output Set: N:\CRF3\04232002\J040895.raw

```
 4 <110> APPLICANT: Sem, Daniel S.
 5         Baker, Brian
 6         Hansen, Mark R.
 8 <120> TITLE OF INVENTION: Methods for Predicting Functional and
 9         Structural Properties of Polypeptides Using Sequence Models
12 <130> FILE REFERENCE: P-TB 5072
14 <140> CURRENT APPLICATION NUMBER: US 10/040,895
C--> 15 <141> CURRENT FILING DATE: 2002-04-09
17 <150> PRIOR APPLICATION NUMBER: US 09/753,020
18 <151> PRIOR FILING DATE: 2000-12-29
20 <160> NUMBER OF SEQ ID NOS: 17
22 <170> SOFTWARE: FastSEQ for Windows Version 4.0
24 <210> SEQ ID NO: 1
25 <211> LENGTH: 155
26 <212> TYPE: PRT
27 <213> ORGANISM: Homo sapiens
29 <400> SEQUENCE: 1
30 Cys Leu Ile Gly Cys Gly Phe Ser Thr Gly Tyr Gly Ala Ala Val Lys
31  .1                5                10                15
32 Thr Gly Lys Val Lys Pro Gly Ser Thr Cys Val Val Phe Gly Leu Gly
33              20                25                30
34 Gly Val Gly Leu Ser Val Ile Met Gly Cys Lys Ser Ala Gly Ala Ser
35              35                40                45
36 Arg Ile Ile Gly Ile Asp Leu Asn Lys Asp Lys Phe Glu Lys Ala Met
37      50                55                60
38 Ala Val Gly Ala Thr Glu Cys Ile Ser Pro Lys Asp Ser Thr Lys Pro
39 65                70                75                80
40 Ile Ser Glu Val Leu Ser Glu Met Thr Gly Asn Asn Val Gly Tyr Thr
41              85                90                95
42 Phe Glu Val Ile Gly His Leu Glu Thr Met Ile Asp Ala Leu Ala Ser
43              100               105               110
44 Cys His Met Asn Tyr Gly Thr Ser Val Val Val Gly Val Pro Pro Ser
45              115               120               125
46 Ala Lys Met Leu Thr Tyr Asp Pro Met Leu Leu Phe Thr Gly Arg Thr
47      130               135               140
48 Trp Lys Gly Cys Val Phe Gly Gly Leu Lys Ser
49 145                   150               155
52 <210> SEQ ID NO: 2
53 <211> LENGTH: 152
54 <212> TYPE: PRT
55 <213> ORGANISM: Equus caballus
57 <400> SEQUENCE: 2
58 Gly Cys Gly Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys
```

RAW SEQUENCE LISTING   DATE: 04/23/2002
PATENT APPLICATION: US/10/040,895   TIME: 15:35:39

Input Set : A:\Tb5072.txt
Output Set: N:\CRF3\04232002\J040895.raw

```
59  1                    5                   10                  15
60  Val Thr Gln Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly
61              20                  25                  30
62  Leu Ser Val Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile
63          35                  40                  45
64  Gly Val Asp Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly
65      50                  55                  60
66  Ala Thr Glu Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu
67  65                  70                  75                  80
68  Val Leu Thr Glu Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val
69              85                  90                  95
70  Ile Gly Arg Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu
71              100                 105                 110
72  Ala Tyr Gly Val Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn
73          115                 120                 125
74  Leu Ser Met Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly
75          130                 135                 140
76  Ala Ile Phe Gly Gly Phe Lys Ser
77  145                 150
80  <210> SEQ ID NO: 3
81  <211> LENGTH: 175
82  <212> TYPE: PRT
83  <213> ORGANISM: Thermoanaerobium Brockii
85  <400> SEQUENCE: 3
86  Val Met Ile Pro Asp Met Met Thr Thr Gly Phe His Gly Ala Glu Leu
87  1                    5                   10                  15
88  Ala Asp Ile Glu Leu Gly Ala Thr Val Ala Val Leu Gly Ile Gly Pro
89              20                  25                  30
90  Val Gly Leu Met Ala Val Ala Gly Ala Lys Leu Arg Gly Ala Gly Arg
91          35                  40                  45
92  Ile Ile Ala Val Gly Ser Arg Pro Val Cys Val Asp Ala Ala Lys Tyr
93      50                  55                  60
94  Tyr Gly Ala Thr Asp Ile Val Asn Tyr Lys Asp Gly Pro Ile Glu Ser
95  65                  70                  75                  80
96  Gln Ile Met Asn Leu Thr Glu Gly Lys Gly Val Asp Ala Ala Ile Ile
97              85                  90                  95
98  Ala Gly Gly Asn Ala Asp Ile Met Ala Thr Ala Val Lys Ile Val Lys
99              100                 105                 110
100 Pro Gly Gly Thr Ile Ala Asn Val Asn Tyr Phe Gly Glu Gly Glu Val
101         115                 120                 125
102 Leu Pro Val Pro Arg Leu Glu Trp Gly Cys Gly Met Ala His Lys Thr
103         130                 135                 140
104 Ile Lys Gly Gly Leu Cys Pro Gly Gly Arg Leu Arg Met Glu Arg Leu
105 145                 150                 155                 160
106 Ile Asp Leu Val Phe Tyr Lys Arg Val Asp Pro Ser Lys Leu Val
107             165                 170                 175
110 <210> SEQ ID NO: 4
111 <211> LENGTH: 141
112 <212> TYPE: PRT
```

RAW SEQUENCE LISTING   DATE: 04/23/2002
PATENT APPLICATION: US/10/040,895   TIME: 15:35:39

Input Set : A:\Tb5072.txt
Output Set: N:\CRF3\04232002\J040895.raw

```
113 <213> ORGANISM: Lactobacillus confusus
115 <400> SEQUENCE: 4
116 Ala Arg Lys Ile Gly Ile Ile Gly Leu Gly Asn Val Gly Ala Ala Val
117 1               5                   10                  15
118 Ala His Gly Leu Ile Ala Gln Gly Val Ala Asp Asp Tyr Val Phe Ile
119             20                  25                  30
120 Asp Ala Asn Glu Ala Lys Val Lys Ala Asp Gln Ile Asp Phe Gln Asp
121             35                  40                  45
122 Ala Met Ala Asn Leu Glu Ala His Gly Asn Ile Val Ile Asn Asp Trp
123     50                  55                  60
124 Ala Ala Leu Ala Asp Ala Asp Val Val Ile Ser Thr Leu Gly Asn Ile
125 65                  70                  75                  80
126 Lys Leu Gln Gln Phe Ala Glu Leu Lys Phe Thr Ser Ser Met Val Gln
127             85                  90                  95
128 Ser Val Gly Thr Asn Leu Lys Glu Ser Gly Phe His Gly Val Leu Val
129             100                 105                 110
130 Val Ile Ser Asn Pro Val Asp Val Ile Thr Ala Leu Phe Gln His Val
131         115                 120                 125
132 Thr Gly Phe Pro Ala His Lys Val Ile Gly Thr Gly Thr
133     130                 135                 140
136 <210> SEQ ID NO: 5
137 <211> LENGTH: 147
138 <212> TYPE: PRT
139 <213> ORGANISM: B. Stearothermophilus
141 <400> SEQUENCE: 5
142 Met Lys Asn Asn Gly Gly Ala Arg Val Val Val Ile Gly Ala Gly Phe
143 1               5                   10                  15
144 Val Gly Ala Ser Tyr Val Phe Ala Leu Met Asn Gln Gly Ile Ala Asp
145             20                  25                  30
146 Glu Ile Val Leu Ile Asp Ala Asn Glu Ser Lys Ala Ile Gly Asp Ala
147         35                  40                  45
148 Met Asp Phe Asn His Gly Lys Val Phe Ala Pro Lys Pro Val Asp Ile
149     50                  55                  60
150 Trp His Gly Asp Tyr Asp Asp Cys Arg Asp Ala Asp Leu Val Val Ile
151 65                  70                  75                  80
152 Cys Ala Gly Ala Asn Gln Lys Pro Gly Glu Thr Arg Leu Asp Leu Val
153             85                  90                  95
154 Asp Lys Asn Ile Ala Ile Phe Arg Ser Ile Val Glu Ser Val Met Ala
155             100                 105                 110
156 Ser Gly Phe Gln Gly Leu Phe Leu Val Ala Thr Asn Pro Val Asp Ile
157         115                 120                 125
158 Leu Thr Tyr Ala Thr Trp Lys Phe Ser Gly Leu Pro His Glu Arg Val
159     130                 135                 140
160 Ile Gly Ser
161 145
164 <210> SEQ ID NO: 6
165 <211> LENGTH: 312
166 <212> TYPE: PRT
167 <213> ORGANISM: E. Coli
```

```
169 <400> SEQUENCE: 6
170 Met Lys Val Ala Val Leu Gly Ala Ala Gly Gly Ile Gly Gln Ala Leu
171 1               5                   10              15
172 Ala Leu Leu Leu Lys Thr Gln Leu Pro Ser Gly Ser Glu Leu Ser Leu
173         20                  25                  30
174 Tyr Asp Ile Ala Pro Val Thr Pro Gly Val Ala Val Asp Leu Ser His
175         35                  40                  45
176 Ile Pro Thr Ala Val Lys Ile Lys Gly Phe Ser Gly Glu Asp Ala Thr
177     50                  55                  60
178 Pro Ala Leu Glu Gly Ala Asp Val Val Leu Ile Ser Ala Gly Val Arg
179 65                  70                  75                  80
180 Arg Lys Pro Gly Met Asp Arg Ser Asp Leu Phe Asn Val Asn Ala Gly
181             85                  90                  95
182 Ile Val Lys Asn Leu Val Gln Gln Val Ala Lys Thr Cys Pro Lys Ala
183             100                 105                 110
184 Cys Ile Gly Ile Ile Thr Asn Pro Val Asn Thr Thr Val Ala Ile Ala
185         115                 120                 125
186 Ala Glu Val Leu Lys Lys Ala Gly Val Tyr Asp Lys Asn Lys Leu Phe
187     130                 135                 140
188 Gly Val Thr Thr Leu Asp Ile Ile Arg Ser Asn Thr Phe Val Ala Glu
189 145                 150                 155                 160
190 Leu Lys Gly Lys Gln Pro Gly Glu Val Glu Val Pro Val Ile Gly Gly
191             165                 170                 175
192 His Ser Gly Val Thr Ile Leu Pro Leu Leu Ser Gln Val Pro Gly Val
193             180                 185                 190
194 Ser Phe Thr Glu Gln Glu Val Ala Asp Leu Thr Lys Arg Ile Gln Asn
195         195                 200                 205
196 Ala Gly Thr Glu Val Val Glu Ala Lys Ala Gly Gly Gly Ser Ala Thr
197     210                 215                 220
198 Leu Ser Met Gly Gln Ala Ala Ala Arg Phe Gly Leu Ser Leu Val Arg
199 225                 230                 235                 240
200 Ala Leu Gln Gly Glu Gln Gly Val Val Glu Cys Ala Tyr Val Glu Gly
201             245                 250                 255
202 Asp Gly Gln Tyr Ala Arg Phe Phe Ser Gln Pro Leu Leu Leu Gly Lys
203             260                 265                 270
204 Asn Gly Val Glu Glu Arg Lys Ser Ile Gly Thr Leu Ser Ala Phe Glu
205         275                 280                 285
206 Gln Asn Ala Leu Glu Gly Met Leu Asp Thr Leu Lys Lys Asp Ile Ala
207     290                 295                 300
208 Leu Gly Gln Glu Phe Val Asn Lys
209 305                 310
212 <210> SEQ ID NO: 7
213 <211> LENGTH: 163
214 <212> TYPE: PRT
215 <213> ORGANISM: Sus scrofa
217 <400> SEQUENCE: 7
218 Ala Thr Leu Lys Asp Gln Leu Ile His Asn Leu Leu Lys Glu Glu His
219 1               5                   10              15
220 Val Pro His Asn Lys Ile Thr Val Val Gly Val Gly Ala Val Gly Met
```

RAW SEQUENCE LISTING   DATE: 04/23/2002
PATENT APPLICATION: US/10/040,895   TIME: 15:35:39

Input Set : A:\Tb5072.txt
Output Set: N:\CRF3\04232002\J040895.raw

```
221           20              25              30
222 Ala Cys Ala Ile Ser Ile Leu Met Lys Glu Leu Ala Asp Glu Ile Ala
223           35              40              45
224 Leu Val Asp Val Met Glu Asp Lys Leu Lys Gly Glu Met Met Asp Leu
225      50              55              60
226 Gln His Gly Ser Leu Phe Leu Arg Thr Pro Lys Ile Val Ser Gly Lys
227 65              70              75              80
228 Asp Tyr Asn Val Thr Ala Asn Ser Arg Leu Val Val Ile Thr Ala Gly
229           85              90              95
230 Ala Arg Gln Gln Glu Gly Glu Ser Arg Leu Asn Leu Val Gln Arg Asn
231           100             105             110
232 Val Asn Ile Phe Lys Phe Ile Ile Pro Asn Ile Val Lys Tyr Ser Pro
233           115             120             125
234 Asn Cys Lys Leu Leu Val Val Ser Asn Pro Val Asp Ile Leu Thr Tyr
235      130             135             140
236 Val Ala Trp Lys Ile Ser Gly Phe Pro Lys Asn Arg Val Ile Gly Ser
237 145             150             155             160
238 Gly Cys Asn
242 <210> SEQ ID NO: 8
243 <211> LENGTH: 333
244 <212> TYPE: PRT
245 <213> ORGANISM: Sus scrofa
247 <400> SEQUENCE: 8
248 Ser Glu Pro Ile Arg Val Leu Val Thr Gly Ala Ala Gly Gln Ile Ala
249 1              5               10              15
250 Tyr Ser Leu Leu Tyr Ser Ile Gly Asn Gly Ser Val Phe Gly Lys Asp
251           20              25              30
252 Gln Pro Ile Ile Leu Val Leu Leu Asp Ile Thr Pro Met Met Gly Val
253           35              40              45
254 Leu Asp Gly Val Leu Met Glu Leu Gln Asp Cys Ala Leu Pro Leu Leu
255      50              55              60
256 Lys Asp Val Ile Ala Thr Asp Lys Glu Glu Ile Ala Phe Lys Asp Leu
257 65              70              75              80
258 Asp Val Ala Ile Leu Val Gly Ser Met Pro Arg Arg Asp Gly Met Glu
259           85              90              95
260 Arg Lys Asp Leu Leu Lys Ala Asn Val Lys Ile Phe Lys Cys Gln Gly
261           100             105             110
262 Ala Ala Leu Asp Lys Tyr Ala Lys Lys Ser Val Lys Val Ile Val Val
263           115             120             125
264 Gly Asn Pro Ala Asn Thr Asn Cys Leu Thr Ala Ser Lys Ser Ala Pro
265      130             135             140
266 Ser Ile Pro Lys Glu Asn Phe Ser Cys Leu Thr Arg Leu Asp His Asn
267 145             150             155             160
268 Arg Ala Lys Ala Gln Ile Ala Leu Lys Leu Gly Val Thr Ser Asp Asp
269           165             170             175
270 Val Lys Asn Val Ile Ile Trp Gly Asn His Ser Ser Thr Gln Tyr Pro
271           180             185             190
272 Asp Val Asn His Ala Lys Val Lys Leu Gln Ala Lys Glu Val Gly Val
273      195             200             205
```

VERIFICATION SUMMARY
PATENT APPLICATION: **US/10/040,895**

DATE: 04/23/2002
TIME: 15:35:40

Input Set : **A:\Tb5072.txt**
Output Set: **N:\CRF3\04232002\J040895.raw**

L:15 M:271 C: Current Filing Date differs, Replaced Current Filing Date